AD-A104 477 JOHNS HOPKINS UNIV BALTIMORE MD DEPT OF MATHEMATICAL--ETC F/G 12/1
ERROR BOUNDS FOR RECONSTRUCTION OF A FUNCTION F FROM A FINITE S--ETC(U)
MAY 81 A F KARR, R J SERFLING N00014-79-C-0801
UNCLASSIFIED TR-341 NL

1 OF 1
AD A
104477

END
DATE
FILMED
10-81
DTIC

AD A104477

(9)

LEVEL II

DEPARTMENT OF MATHEMATICAL SCIENCES
The Johns Hopkins University
Baltimore, Maryland 21218

(6) ERROR BOUNDS FOR RECONSTRUCTION OF A FUNCTION f FROM A FINITE SEQUENCE $\{SGN(f(t_i) + x_i)\}$

by

(12) 20

(10) Alan F. Karr ~~and~~ Robert J. Serfling
The Johns Hopkins University

(9) Technical Report, No. 341
ONR Technical ~~Report~~ No. 81-4

(11) May 1981

(14) TR-341, 81-4-ONR

(15)

Research supported by the Army, Navy and Air Force under Office of Naval Research Contract No. N00014-79-C-0801.
Reproduction in whole or in part is permitted for any purpose of the United State Government.

408328

81 9 22 040

ABSTRACT

ERROR BOUNDS FOR RECONSTRUCTION OF A FUNCTION f FROM A FINITE SEQUENCE $\{SGN(f(t_i) + x_i)\}$

Consider reconstructing a function $f(t)$, $0 \le t \le 1$, from knowledge only of $\{(t_i, s_i), 1 \le i \le n\}$, where $s_i = \mathrm{sgn}(f(t_i) + x_i)$, $1 \le i \le n$, and the x_i are additive "corruptions." Without the components x_i, f could not be reconstructed. However, for f continuous and for random uniform noise x, Masry and Cambanis (1980,1981) show that f can be consistently estimated almost surely and in mean square as $n\to\infty$ and establish rates of uniform convergence. Through a somewhat different treatment, in which the approximation of $f(t_i)$ is identified as a numerical integration problem rather than a statistical problem, we obtain simple exact bounds on the error of estimation, allow the noise x to be arbitrary (random or deterministic), and deal with the case of f having discontinuities. The bounds yield substantially improved rates of convergence when the noise values x_i are from a quasi-random instead of random sequence.

1. Introduction. Consider reconstructing a function $f(t)$, $0 \le t \le 1$, from knowledge only of $\{(t_i, s_i), 1 \le i \le n\}$, where $s_i = \mathrm{sgn}(f(t_i) + x_i)$, $1 \le i \le n$, and the x_i are additive "corruptions" whose values may or may not be known. In the absence of the x_i, the resulting values s_i permit only the estimation of sgn $f(t)$, $0 \le t \le 1$, from which, of course, f cannot be reconstructed. However, with appropriate $\{x_i\}$, and under minimal assumptions on f, the given "data" permits approximation of f within any desired accuracy with respect to the L^p-metrics, $1 \le p \le \infty$, as $n\to\infty$. In this paper we develop suitable approximating functions $\hat{f}(t)$, $0 \le t \le 1$, and establish explicit and useful upper bounds on the approximation error. We also derive convergence rates under relevant conditions.

Such results have application, for example, to communication systems. An unknown continuous-time real signal $f(t)$ is to be reduced to binary form ("hardlimited"), transmitted, then reconstructed. In this context of nonparametric signal identification, the problem has been previously considered by Masry and Cambanis (1980, 1981) and Masry (1981). They assume that f is continuous and is bounded in magnitude by a known constant B and show that if the sequence $f(t_i)$ is deliberately corrupted additively by a sequence of uniform $[-B,B]$ random variables X_i before hardlimiting, then f can be estimated consistently almost surely and in mean square as $n\to\infty$. They also establish rates for these convergences.

The approach of Masry and Cambanis derives from the fact that, for any t_i, $f(t_i)$ may be represented as an expectation:

$$f(t_i) = E[B\ \mathrm{sgn}(f(t_i) + X_i)],$$

so that an unbiased estimator of $f(t_i)$ is given by

$$\hat{f}_0(t_i) = B\ \mathrm{sgn}(f(t_i) + X_i),\ 1 \le i \le n.$$

They construct $\hat{f}_0(t)$ elsewhere by interpolation. In their treatment the contribution to the approximation error $|\hat{f}_0(t) - f(t)|$ from interpolation error is greatly dominated by the contribution due to stochastic variation in the estimates $\hat{f}_0(t_i)$. It is important, therefore, to seek a reduction of the "stochastic error" component.

In the present paper the approximation of $f(t_i)$ is identified as a numerical integration problem rather than a statistical problem. In this context, the method of Masry and Cambanis corresponds to a Monte Carlo approach. Viewing the problem in this fashion, we are able to allow any kind of noise sequence $\{x_i\}$, random or deterministic, and to obtain general exact upper bounds on the approximation error. In particular, exploiting modern improvements in the Monte Carlo method, we replace the random sequence $\{X_i\}$ by a suitable "quasi-random" sequence $\{x_i\}$. This permits the stochastic error to be replaced by a much smaller "numerical integration error" counterpart, leading to radical improvements in the rates of convergence in the metrics of interest. The approach also handles the case of f having discontinuities and can be extended to f defined on a multi-dimensional domain.

A precise formulation of the problem is presented in Section 2

and a suitable approximating function $\hat{f}$ is introduced. The approach described above is implemented in Section 3 to obtain bounds on the L^∞-metric approximation error $\|\hat{f} - f\|_\infty$, in terms of the <u>modulus of continuity</u> of the function f and the <u>discrepancy</u> (from exact uniformity) of the sequence $\{x_i\}$. For example, for f Lipschitz of order 1 (Lip 1) and for suitably chosen $\{x_i\}$, our bounds yield the rate $O(n^{-1/2})$ for the convergence of $\|\hat{f} - f\|_\infty$ to 0 as $n \to \infty$, where n is the number of evaluations of $\mathrm{sgn}(f(t_i) + x_i)$. The number n is an appropriate measure of the "work" involved in calculating the estimator. In comparison, Masry and Cambanis (1981) obtain the rate $O(n^{-1/4+\epsilon})$, $\epsilon>0$, for $f \in$ Lip 1 and $\{X_i\}$ random, with a nonsequential estimator and, for a sequential estimator of moving average form, Masry (1981) obtains the rate $O(n^{-1/3+\epsilon})$, $\epsilon>0$. Our bounds are applicable to random $\{X_i\}$ and yield the sharper rate $O((\log n)^{1/3} n^{-1/3})$. Moreover, our estimator is simpler and easier to compute that the moving average estimator of Masry (1981).

We also deal with the case of f having discontinuities, for which purpose we consider the L^p-metrics, $1 \le p < \infty$. Using the above ideas, bounds for $\|\hat{f} - f\|_p$ are obtained in Section 4, for f of bounded variation. For suitable choice of $\{t_i\}$ and $\{x_i\}$, the rate $O((\log n)^{1/2} n^{-1/2(p+1)})$ follows.

There are other ways to use data of the specific form $\{(t_i, s_i)\}$ to determine the approximate whereabouts of the function f. Two alternative approaches leading to useful bounds on the L^∞-metric

approximation error are presented in Section 5. However, the numerical integration approach extends to multi-dimensional settings, as discussed in Section 6 along with other comments.

2. Formulation of the problem. Let $f(t)$, $0 \leq t \leq 1$, be an unknown function satisfying

(2.1) $$|f(t)| \leq B, \; 0 \leq t \leq 1,$$

with B known and finite. Define

$$q(y,u) = B \operatorname{sgn}(y + 2Bu - B), \; |y| \leq B, \; 0 \leq u \leq 1.$$

It is easily checked that $\int_0^1 q(y,u)du = y$, $|y| \leq B$. Hence we may represent $f(t)$ as

(2.2) $$f(t) = \int_0^1 q(f(t), u)du.$$

Our objective is to construct a suitable approximating function $\hat{f}(t)$, $0 \leq t \leq 1$, on the basis of "data" of the form $\{(t_i,s_i),\ 1 \leq i \leq n\}$, where $s_i = q(f(t_i), u_i)$, with $\{u_i\}$ a sequence in $[0,1]$, and to give suitable bounds on the error of approximation. A straightforward approach is to estimate $f(t)$ directly at J selected points $t = t_1', \ldots, t_J'$ and obtain $\hat{f}(t)$ elsewhere by interpolation or by a step function. By (2.2), the estimation of f at a selected point t' may be viewed as a problem of evaluating an integral, for which we would want to have evaluation of the integrand $q(t', u)$ at K suitable points $u = u_1', \ldots, u_K'$. In this case the desired

data form an array $\{(t_j', s_{jk}'),\ 1 \leq j \leq J,\ 1 \leq k \leq K\}$, with $s_{jk}' = q(f(t_j'), u_{jk}')$. However, a further complication must be taken into account: in practice, since the data typically is collected in "real time", $q(f(t), u)$ might not be observable for more than one value of u per value of t. To allow for this possibility, we shall assume that the data is an array of the form

(2.3a) $$\{(t_{jk}, s_{jk}),\ 1 \leq j \leq J,\ 1 \leq k \leq K\},$$

where

(2.3b) $$0 \leq t_{11} \leq \cdots \leq t_{1K} < t_{21} \leq \cdots \leq t_{J-1,K} < t_{J1} \leq \cdots \leq t_{JK} \leq 1,$$

and where $s_{jk} = q(f(t_{jk}), u_{jk})$. When the inequalities in (2.3b) are strict, the estimation of f at any specified point is more difficult.

In addition, the double array format (2.3) presents us with a design problem, of choosing the best trade-off between J and K for a fixed choice of the number $n = JK$ of evaluations of $\operatorname{sgn}(f(t) + x)$. As will be seen, this can be viewed as a trade-off between a numerical integration error and an interpolation error.

We confine attention to step function estimators. This does not increase the order of magnitude of the interpolation error in comparison with linear interpolation, for example, and has the

advantages of simplicity and computational ease. Specifically, we assume that $[0,1]$ is divided into intervals $I_j = [t_{j-1}, t_j)$, $j = 1, \ldots, J$, where $0 = t_0 < t_1 < \ldots < t_J = 1$ and that the estimator $\hat{f}(t)$, $0 \le t \le 1$, satisfies

$$\text{(2.4)} \qquad \hat{f}(t) \text{ is constant on each interval } I_j.$$

Moreover, the above t_j and the t_{jk} of (2.3) are selected so that $t_{jk} \in I_j$ for each j and k, and the value of $\hat{f}$ on I_j is a function only of the relevant portion of the data (2.3), namely $\{(t_{jk}, s_{jk})$, $1 \le k \le K\}$. The determination of $\hat{f}(t)$ within I_j will now be described.

One natural approach is based on the idea of estimating the integral in (2.2) by a suitable average of values of q. For the data (2.3) this idea leads to the estimator

$$\text{(2.5)} \qquad \hat{f}(t) = \frac{1}{K}\sum_{k=1}^{K} q(f(t_{jk}), u_{jk}), \quad t \in I_j,\ 1 \le j \le J.$$

We consider this estimator in Sections 3 and 4 and establish bounds on $\|\hat{f} - f\|_\infty$ and $\|\hat{f} - f\|_p$ $(1 \le p < \infty)$, respectively, as discussed in the Introduction.

The data (2.3) may be used in other ways to develop an approximation to a continuous f. Alternatives to the estimator (2.5) are treated in Section 5.

The estimator (2.5) is very efficient computationally. Over each interval I_j it is necessary to maintain *only a running total*

$K^{-1}\sum_{1}^{\ell} q(f(t_{jk}), u_{jk})$, which becomes $\hat{f}(t)$, $t \in I_j$, when ℓ reaches K. On the other hand, the moving average estimator of Masry (1981), whose convergence properties are compared in Section 3 to those of the estimator (2.5), is less efficient in that it entails more computations and larger storage requirements. In our notation the estimator of Masry (1981) is a moving average of the form

$$\text{(2.6)} \qquad \tilde{f}(t_\ell') = \frac{1}{p}\sum_{k=1}^{p} q(t_{\ell-k}', U_{\ell-k})$$

at selected points $t_1', \ldots, t_n'$ in $[0,1]$, where the U_i are independent uniform $[0,1]$ random variables, with $\tilde{f}$ defined elsewhere by linear interpolation. To obtain the estimator at the t_ℓ' requires calculation and storage of n averages of the form (2.6). For the estimator (2.5), JK is the number of function evaluations, and is comparable to n above, but the number of averages calculated and stored is J.

3. <u>Approximation in L^∞</u>. Consider the problem as formulated in Section 2. For approximation of a function f satisfying (2.1) and (2.2) by the function $\hat{f}$ given by (2.4) and (2.5), we will derive an upper bound on $\|\hat{f} - f\|_\infty = \sup_t |\hat{f}(t) - f(t)|$. The bound will involve the <u>modulus of continuity</u> of f, $\omega(f;\delta) = \sup_{|s-t| \le \delta} |f(s) - f(t)|$, and the <u>discrepancy</u> of the values $\{u_{jk}\}$. For a sequence $\langle u \rangle = u_1, u_2, \ldots$ in $[0,1)$, the <u>discrepancy</u> of the initial

segment of length K is defined as

$$D_K^*(\langle u\rangle) = \sup_{0<v<1} \left| \frac{A([0,v);\, K;\, \langle u\rangle)}{K} - v \right|,$$

where $A(E;\, K;\, \langle u\rangle)$ denotes the number of $u_1, \ldots, u_K$ which belong to the set E. This quantity measures the departure of $u_1, \ldots, u_K$ from a "uniform" sequence (see Kuipers and Niederreiter (1974) and Niederreiter (1978) for excellent expository discussion of discrepancy and related topics). We now can state the main result of this section.

THEOREM 3.1. *Assume that* $|f(t)| \le B$, $0 \le t \le 1$, *where* B *is known, and that the data* (2.3) *is available. Let* $\hat{f}(t)$, $0 \le t \le 1$, *be given by* (2.4) *and* (2.5). *Then*

$$\|\hat{f} - f\|_\infty \le 2B \max_{1\le j\le J} D_K^*(\langle u_j\rangle) + \max_{1\le j\le J} \omega(f; t_j - t_{j-1}), \tag{3.1}$$

where $D_K^*(\langle u_j\rangle)$ *is the discrepancy of the sequence* $u_{j1}, \ldots, u_{jK}$.

The proof of the theorem will make use of an elementary inequality due to Koksma (1942) (or see Niederreiter (1978)), which arises in connection with the approximation of an integral of the form $\int_0^1 g(u)du$ by averages of the form $(1/K)\, \Sigma_1^K g(u_k)$, where $0 \le u_1, \ldots, u_K \le 1$. The approximation error is bounded as follows.

LEMMA 3.1. *If* g *has finite variation* V(g) *on* [0,1] *and* $\langle u\rangle = u_1, u_2, \ldots$ *is any sequence in* [0,1], *then for each* K

$$\left| \int_0^1 g(u)du - \frac{1}{K}\sum_{k=1}^K g(u_k) \right| \le V(g)\, D_K^*(\langle u\rangle). \tag{3.2}$$

PROOF OF THEOREM 3.1. Although the integral in (2.2) is of the form $\int_0^1 g(u)du$, we cannot directly apply (3.2) because the average in (2.5) is not of the form $(1/K)\, \Sigma_1^K g(u_k)$. However, since $q(y,u)$ is nondecreasing in y for each fixed u, we have

$$\frac{1}{K}\sum_{k=1}^K q(m_j, u_{jk}) \le \hat{f}(t) \le \frac{1}{K}\sum_{k=1}^K q(M_j, u_{jk}),\ t \in I_j, \tag{3.3}$$

where m_j and M_j are the infimum and supremum, respectively, of f over the interval I_j. Applying (3.2), noting that $V(q(y,\cdot)) = 2B$ for all y and recalling that $\int_0^1 q(y,u)du = y$, we obtain

$$\left|\frac{1}{K}\sum_{k=1}^K q(m_j, u_{jk}) - m_j\right| \le 2B\, D_K^*(\langle u_j\rangle) \tag{3.4}$$

and

$$\left|\frac{1}{K}\sum_{k=1}^K q(M_j, u_{jk}) - M_j\right| \le 2B\, D_K^*(\langle u_j\rangle). \tag{3.5}$$

By a simple argument using (3.3) - (3.5) and the relation $m_j \le f(t) \le M_j$, $t \in I_j$, it follows that for $t \in I_j$

$$|\hat{f}(t) - f(t)| \le 2B\, D_K^*(\langle u_j\rangle) + M_j - m_j,$$

which yields (3.1). //

For J fixed, the modulus of continuity term in (3.1) is minimized by taking the intervals I_j all of length 1/J. Likewise, for K fixed, the discrepancy term attains its minimum possible value 1/2K when the u_{jk}, $1 \le k \le K$, are taken equally spaced in [0,1] for each j, that is, for $u_{jk} = (2k-1)/(2K)$, $1 \le k \le K$,

$1 \le j \le J$. With these substitutions, the inequality (3.1) becomes

(3.6) $$\|\hat{f} - f\|_\infty \le \frac{B}{K} + \omega(f; \frac{1}{J}).$$

Suppose, further, that $f \in \mathrm{Lip}\ \gamma$ on $[0,1]$. Then (3.6) may be replaced by

(3.7) $$\|\hat{f} - f\|_\infty \le C(K^{-1} + J^{-\gamma}),$$

where C is a constant. Finally, suppose that we choose J and K to maximize the rate of convergence of this bound to 0 as $n\to\infty$. We obtain

COROLLARY 3.1. Assume that $|f| \le B$, that $f \in \mathrm{Lip}\ \gamma$, $0 < \gamma \le 1$, on $[0,1]$, and that $\hat{f}$ is given by (2.5) with $J = J_n = n^{1/(\gamma+1)}$, $t_j = j/J_n$, $K = K_n = n^{\gamma/(\gamma+1)}$, and $u_{jk} = (2k-1)/2K_n$. Then

(3.8) $$\|\hat{f} - f\|_\infty = O(n^{-\gamma/(\gamma+1)}), \qquad n\to\infty.$$

It should be noted that (3.3) and (3.6) are applicable to discontinuous f, whereas the results of Masry and Cambanis (1981) and Masry (1981) apply only to continuous f. Of course the bound in (3.6) does not go to zero as $J, K\to\infty$ if f is discontinuous but nevertheless does assure that f can be reconstructed to within an error $\omega(f;\ 0) = \lim_{\delta\to 0} \omega(f, \delta)$.

Returning to the problem of choosing the u_{jk}, in some cases it is desirable to let the u_{jk}, $1 \le k \le K$, be generated by a single sequence $<u>$ rather than recompute the values for each change of K.

It is known that the *fastest rate* possible for convergence of $D_K^*(<u>)$ to 0, as $K\to\infty$ with $<u>$ an infinite sequence is $O((\log K)/K)$. This rate is attained by the van der Corput sequence $<v>$ defined as follows:

$$v_k = \sum_{j=0}^{s} a_j 2^{-(j+1)},$$

where s and $a_0, \ldots, a_s$ are determined by

$$k = \sum_{j=0}^{s} a_j 2^j.$$

In particular, the discrepancy of the van der Corput sequence satisfies

$$D_K^*(<v>) \le \frac{\log (K+1)}{(\log 2)\ K}, \qquad K \ge 1.$$

See Kuipers and Niederreiter (1974) for details concerning the van der Corput sequence and other low discrepancy sequences. With use of the van der Corput sequence, a minor relaxation of the convergence rate in (3.8) results.

However, the rate in (3.8) breaks down radically if the u_{jk} are replaced by independent *uniform* $[0,1]$ random variables $\{U_{jk}\}$. This is because for such a sequence $<U>$, the quantity $D_K^*(<U>)$ is precisely the Kolmogorov-Smirnov test statistic, which converges almost surely to 0 at the exact rate $O((\log\log K)^{1/2} K^{-1/2})$ as $K\to\infty$ (cf. Chung (1949)). This illustrates the limitations of the Monte

Carlo method in comparison with the so-called quasi-Monte Carlo method based on quasi-random sequences such as the van der Corput sequence (see Niederreiter (1978) for discussion). To derive the counterpart to (3.8) when the u_{jk} are random uniform variables instead of equally spaced values, we shall apply an inequality of Dvoretsky, Kiefer and Wolfowitz (1956), which may be stated as follows.

LEMMA 3.2. Let $\langle U\rangle$ be a sequence of independent uniform [0,1] random variables. Then for each K

$$(3.9) \qquad P\{D_K^*(\langle U\rangle) > d\} \le Ce^{-2Kd^2}, \qquad d>0,$$

where C is a universal constant not depending on K.

We then obtain the following.

PROPOSITION 3.1. Assume that $|f|\le B$ and $f \in \mathrm{Lip}\,\gamma$ on $[0,1]$, and let

$$(3.10) \qquad \hat f(t) = \frac{1}{K}\sum_{k=1}^{K} q(f(t_{jk}), U_{jk}), \qquad t \in I_j,$$

where $J = J_n = n^{1/(2\gamma+1)}(\log n)^{-1/(2\gamma+1)}$, the I_j all have length $1/J_n$, $K = K_n = n^{2\gamma/(2\gamma+1)}(\log n)^{1/(2\gamma+1)}$, and where the U_{jk} are independent, uniform [0,1] random variables. Then almost surely

$$(3.11) \qquad \|\hat f - f\|_\infty = O((\log n)^{\gamma/(2\gamma+1)} n^{-\gamma/(2\gamma+1)}), \qquad n\to\infty.$$

PROOF. It follows using (3.9) that the (now random) discrepancy term in (3.1) satisfies

$$(3.12) \qquad P\{\max_{1\le j\le J} D_K^*(\langle U_j\rangle)>d\} \le JCe^{-2Kd^2}, \qquad d > 0,$$

where $D_K^*(\langle U_j\rangle)$ is the discrepancy of the sequence $U_{j1}, \ldots, U_{jK}$. Let us for the moment take $J_n = n^{\alpha}(\log n)^{-\beta}$, $K_n = n^{1-\alpha}(\log n)^{\beta}$, and $d_n = A(\log n)^{(1-\beta)/2} n^{-(1-\alpha)/2}$, where $0<\alpha<1$, $\beta>0$ and $2A^2>\alpha + 1$.

Then (3.12) yields

$$\sum_{n=1}^{\infty} P\{\max_{1\le j\le J_n} D_{K_n}^*(\langle U_j\rangle)>d_n\} = O(\sum_{n=1}^{\infty} n^{\alpha-2A^2})<\infty.$$

Therefore, by the Borel-Cantelli lemma,

$$(3.13) \qquad \max_{1\le j\le J_n} D_{K_n}^*(\langle U_j\rangle) = O(d_n), \qquad n \to \infty,$$

almost surely. Application of (3.1) and (3.13) with the specified values for J_n, K_n and d_n yields (3.11). //

For $\gamma = 1$ the rate in (3.11) is $O((\log n)^{1/3} n^{-1/3})$, which slightly sharpens the rate $O(n^{-1/3+\epsilon})$, $\epsilon>0$, obtained by Masry (1981). However, the corresponding rate $O(n^{-1/2})$ in (3.8) represents a

dramatic improvement. Further improvement results if the u_{jk} are chosen adaptively; see Section 5.

4. Estimation in L^p $(1 \le p < \infty)$. In this section we investigate the properties of the estimator $\hat{f}$ given by (2.5) with respect to the L^p-norms

$$\|f - g\|_p = \left(\int_0^1 |f(t) - g(t)|^p dt\right)^{1/p}$$

for $1 \le p < \infty$. The importance of doing so is that we are able to provide rates for the convergence $\|\hat{f} - f\|_p \to 0$ for a wide class of discontinuous functions f. This version of the problem, despite its clear significance, seems to have received virtually no previous attention. However, the numerical integration formulation of the problem is extremely natural in this case, so that our methods provide a unified approach to the discontinuous and continuous cases, as well as to the L^∞ and L^p $(1 \le p < \infty)$ cases.

We again deal with the estimator $\hat{f}$ given by (2.5); however, in the present exposition it is convenient to represent the points $t_{jk} \in I_j = [t_{j-1}, t_j)$ in the form

$$t_{jk} = t_{j-1} + v_{jk}(t_j - t_{j-1}),$$

where $0 \le v_{jk} \le 1$. Then

$$\hat{f}(t) = \frac{1}{K}\sum_{k=1}^{K} q(f(t_j + v_{jk}(t_j - t_{j-1})), u_{jk}), \quad t \in I_j. \tag{4.1}$$

The data (2.3) are now represented as $\{(v_{jk}, s_{jk}):\ 1 \le j \le J,$

$1 \le k \le K\}$, where $s_{jk} = q(f(t_{j-1} + v_{jk}(t_j - t_{j-1})), u_{jk})$.

The main result of this section provides an estimate for $\|\hat{f} - f\|_p$, $1 \le p < \infty$, in terms of the following notion of two-dimensional discrepancy.

For a two-dimensional sequence $\langle v,u\rangle$ in $[0,1]^2$, the discrepancy $D_K^*(\langle v,u\rangle)$ is given by

$$D_K^*(\langle v,u\rangle) = \sup_{I^*}\left|\frac{A(I^*;\ K;\ \langle v,u\rangle)}{K} - \lambda(I^*)\right|,$$

where the supremum is over all intervals $I^* = [0,y) \times [0,z)$ and λ denotes Lebesgue measure; see Kuipers and Niederreiter (1974) or Niederreiter (1978) for further details.

THEOREM 4.1 Assume that $|f| \le B$ and that f is of bounded variation on [0,1], and let $\hat{f}$ be given by (4.1). Then for each $p \in [1, \infty)$,

$$\|\hat{f} - f\|_p \le C_1 \max_{1\le j\le J} D_K^*(\langle v_j,u_j\rangle)^{1/2} + C_2 [\max_{1\le j\le J} (t_j - t_{j-1})]^{1/2p}, \tag{4.2}$$

where $D_K^*(\langle v_j, u_j\rangle)$ is the two dimensional discrepancy of the sequence $(v_{j1}, u_{j1}), \ldots, (v_{jK}, u_{jK})$, C_1 is a constant depending on f only through its variation V(f) and the bound B, and C_2 is a constant depending on p, V(f) and B. (These constants are cal-

culated explicitly in the proof.)

PROOF. Define

$$\tilde{f}(t) = \Delta_j^{-1} \int_{I_j} f(v)\, dv, \qquad t \in I_j, \tag{4.3}$$

for $j = 1, \ldots, J$, where $\Delta_j = t_j - t_{j-1} = \lambda(I_j)$. In probabilistic terminology, the function $\tilde{f}$ is the conditional expectation of f (in the probability space $([0,1], B[0,1], \lambda)$) given the σ-algebra generated by the intervals I_j. This implies, for example, that $\tilde{f}$ is the best approximation to f, in the L^2-norm, among functions constant over each I_j. Let $\hat{f}_j$ and $\tilde{f}_j$ be the values of $\hat{f}$ and $\tilde{f}$, respectively, on the interval I_j. We then have

$$\begin{aligned} \|\hat{f} - f\|_p &\le \|\hat{f} - \tilde{f}\|_p + \|\tilde{f} - f\|_p \\ &\le \|\hat{f} - \tilde{f}\|_\infty + \|\tilde{f} - f\|_p \\ &= \max_{1 \le j \le J} |\hat{f}_j - \tilde{f}_j| \quad \|\tilde{f} - f\|_p \end{aligned} \tag{4.4}$$

In this inequality the first term can be regarded as a numerical integration error and the second as an interpolation error (or a smoothing error.)

Now fix j. The term "numerical integration error" is justified by the observation that

$$\tilde{f}_j = \iint_{[0,1]^2} q(f(t_{j-1} + v\Delta_j), u)\, du.$$

In particular, we then have

$$\begin{aligned} |\hat{f}_j - \tilde{f}_j| &= \left| \frac{1}{K}\sum_{k=1}^{K} q(f(t_{j-1} + v_{jk}\Delta_j), u_{jk}) - \iint_{[0,1]^2} q(f(t_{j-1} + v\Delta_j), u)\, dv\, du \right| \\ &= 2 \left| \frac{1}{K}\sum_{k=1}^{K} 1_{E_j}(v_{jk}, u_{jk}) - \iint_{[0,1]^2} 1_{E_j}(v,u)\, dv\, du \right|, \end{aligned} \tag{4.5}$$

where 1_E denotes the indicator function of the set E and

$$E_j = \{(v, u):\ u \le (2B)^{-1}[B - f(t_{j-1} + v\Delta_j)]\}.$$

The multidimensional version of the Koksma inequality (3.2) - see Kuipers and Niederreiter (1974) - is not applicable because the function $(v,u) \to q(f(t_{j-1} + v\Delta_j), u)$ is not of bounded variation in the sense of Hardy and Krause. However, since f is of bounded variation on $[0,1]$, the set E_j belongs to the class M_b, in the context and notation of Niederreiter (1978, pp. 968 - 969), for

$$b(\varepsilon) = (1 + V_j)\varepsilon,$$

where V_j is the variation of the function $v \to (2B)^{-1}[B - f(t_{j-1} + v\Delta_j)]$, which satisfies

$$V_j \le (2B)^{-1} V(f). \tag{4.6}$$

By Theorems 2.11 and 3.1 and expression (3.1) of Niederreiter (1978),

$$\left| \frac{1}{K}\sum_{k=1}^{K} 1_{E_j}(v_{jk},u_{jk}) - \iint_{[0,1]^2} 1_{E_j}(v,u)\,dv\,du \right| \le D_K(M_c) \tag{4.7}$$

$$\le 2[4\cdot 2^{1/2}\gamma_j + 2\gamma_j + 1]\,(D_K^*(<v_j, u_j>))^{1/2},$$

where $c(\varepsilon) = b(\varepsilon) + 4\varepsilon = (5 + V_j)\varepsilon$, $D_K(M_c)$ is the discrepancy associated with the class M_c and $\gamma_j = 5 + V_j$. Combining (4.5)-(4.7), we obtain

$$|\hat{f}_j - \bar{f}_j| \le 4[2\cdot 2^{1/2}\frac{V(f)}{B} + 20\cdot 2^{1/2} + 11]\,[D_K^*(<v_j,u_j>)]^{1/2}, \tag{4.8}$$

which establishes the first term of the bound (4.2) with C_1 the constant on the right-hand side of (4.8).

Finally, we consider the interpolation (smoothing) error $\|\tilde{f} - f\|_p$. By expression (11) of Berens and DeVore (1976),

$$\|\tilde{f} - f\|_p \le C'\{\delta^2 B + \omega_{2,p}(f;\delta)\}, \tag{4.9}$$

where C' is a constant depending only on p, $\omega_{2,p}$ is the second-order L^p modulus of smoothness of f (see Berens and DeVore (1976) or Timan (1963)), and $\delta = \|\tilde{e}_1 - e_1\|_p$ for $e_1(x) = x$. By straightforward calculations (see Example 1 of Berens and DeVore (1976)) one verifies that

$$\delta^2 \le \frac{1}{2}(p+1)^{-1/p} \max_{1\le j\le J} (t_j - t_{j-1}), \tag{4.10}$$

while the results from Timan (1963), p. 127, imply that

$$\omega_{2,p}(f;\varepsilon) \le (2^{1+1/p}V(f))\varepsilon^{1/p}, \tag{4.11}$$

where $V(f)$ is the variation of f over $[0,1]$. Consequently, by (4.9)-(4.11),

$$\|\tilde{f} - f\|_p \le C'\{\frac{B}{2}(p+1)^{-1/p} \max_{1\le j\le J}(t_j - t_{j-1})$$

$$+ 2^{1/2p}V(f)(p+1)^{-1/2p^2}(\max_{1\le j\le J}(t_j - t_{j-1}))^{1/2p}\}$$

$$\le C_2[\max_{1\le j\le J}(t_j - t_{j-1})]^{1/2p},$$

where $C_2 = 2C' \max\{\frac{B}{2}(p+1)^{-1/p}, 2^{1/2p}V(f)(p+1)^{-1/2p^2}\}$,

which provides the second term in (4.2) and completes the proof of the Theorem. //

For fixed J the second term in (4.2) is minimized by taking $t_j = j/J$ (i.e., taking the intervals I_j of equal length $1/J$). By suitable choice of the sequences $\langle v_j, u_j \rangle$ as initial segments of infinite sequences - see Theorem 3.6 of Niederreiter (1978) - one can realize the (order-of-magnitude) lower bound

$$D_K^*(\langle v_j, u_j \rangle) = O\left(\frac{\log K}{K}\right), \qquad K \to \infty, \tag{4.12}$$

for each j. (Two suitable sequences are the Hammersley sequence and the Halton sequence.) Therefore, the bound (4.2) assumes the form

$$\|\hat{f} - f\|_p \le C\left[\left(\frac{\log K}{K}\right)^{1/2} + J^{-1/2p}\right],$$

where C is a constant.

For $n = JK \to \infty$ and an appropriate allocation of effort between numerical integration and interpolation, we obtain the following result.

COROLLARY 4.1. <u>Let</u> $J_n = n^{p/(p+1)}$ <u>with the</u> I_j <u>all of length</u> $1/J_n$ <u>and let</u> $K_n = n^{1/(p+1)}$ <u>with the sequences</u> $\langle v_j, u_j \rangle$ <u>chosen to satisfy</u> (4.12). <u>Then</u>

$$\|\hat{f} - f\|_p = O((\log n)^{1/2} n^{-1/(2p+1)}), \qquad n \to \infty. \tag{4.13}$$

When f is continuous the inequality

$$\|\hat{f} - f\|_p \le \|\hat{f} - f\|_\infty$$

together with (3.2), (3.7) or (3.8) sometimes, but not always, yields a better estimate than those derived in this section. For example, for $f \in \text{Lip } 1$ and $p = 1$, the rate of convergence in (3.8) is $O(n^{-1/2})$, compared to the rate $O((\log n)^{1/2} n^{-1/3})$ in (4.13). However, as f becomes less smooth, (3.8) deteriorates whereas (4.13) does not. If $f \in \text{Lip } 1/4$, then the rate in (3.8) is $O(n^{-1/5})$, but the rate (4.13) remains $O\left[(\log n)^{1/2} n^{-1/3}\right]$. Also in comparison with Theorem 3.1, we note that the bound in Theorem 4.1 depends rather strongly on the distribution of the $t_{jk} = t_{j-1} + v_{jk}(t_j - t_{j-1})$ within the interval I_j, whereas the bound in Theorem 3.1 exhibits no dependence on the distribution of the t_{jk}.

As in Section 3 we may replace the quasi-random numbers u_{jk} by independent, uniform [0,1] random variables U_{jk} and obtain an almost sure rate for the convergence of $\|\hat{f} - f\|_p$ to zero. The estimator below incorporates an efficient division of labor between numerical integration and interpolation for specified $n(= JK)$; this division depends on p.

PROPOSITION 4.1. <u>Assume that</u> $|f| \le B$ <u>and that</u> f <u>is of bounded variation on</u> [0,1]. <u>Let</u> $p \in [1,\infty)$ <u>be fixed and for each</u> n <u>let</u>

$$\hat{f}_n(t) = \frac{1}{K_n} \sum_{k=1}^{K_n} q\left(f\left(\frac{j+v_{jk}}{J_n}\right), U_{jk}\right), \qquad t \in I_j, \tag{4.14}$$

where $I_j = [(j-1)/J_n,\ j/J_n)$, $J = J_n = n^{p/(p+2)}$, $K = K_n = n^{2/(p+2)}$, where $v_{jk} = (2k-1)/2K$ for each j and k, and where the U_{jk} are independent, uniform $[0,1]$ random variables. Then almost surely

$$(4.15) \qquad \|\hat{f}_n - f\|_p = O((\log n)^{1/4} n^{-1/2(p+2)}), \qquad n\to\infty.$$

Note that the v_{jk} of (4.1) remain nonrandom and, in fact, are taken to be the minimum discrepancy sequence of length K.

The proof of Proposition 4.1 is based on a multidimensional analogue of Lemma 3.2, due to Kiefer (1961).

LEMMA 4.1. Let $\langle V, U\rangle$ be a sequence of independent random variables, with each (V_k, U_k) uniformly distributed on $[0,1]^2$. Then for each $\varepsilon>0$ there is a constant C (depending on ε but not on K) such that

$$(4.16) \qquad P\{D_K^*(\langle V, U\rangle)>d\} \le Ce^{-(2-\varepsilon)Kd^2}$$

for all $d>0$.

PROOF OF PROPOSITION 4.1. To begin, we observe that

$$(4.17) \qquad P\{\max_{1\le j\le J_n} D_{K_n}^*(\langle v_j, U_j\rangle)^{1/2}>d\} \le J_n P\{D_K^*(\langle v_1, U_1\rangle)>d^2\}.$$

If the V_{1k} are independent, uniform $[0,1]$ random variables independent of the U_{1k}, then since $\langle v_1\rangle$ is the minimum discrepancy sequence of length K_n we have

$$(4.18) \qquad P\{D_{K_n}^*(\langle v_1, U_1\rangle)>d^2\} \le P\{D_{K_n}^*(\langle V_1, U_1\rangle)>d^2\}$$

by conditioning on $\langle U_1\rangle$. Now choose $\varepsilon>0$ sufficiently small that $\varepsilon<1-p/(p+2)$, and let

$$d_n = (\log n)^{1/4} K_n^{-1/4}.$$

Applying (4.16) - (4.18) we infer that

$$\sum_n P\{\max_{1\le j\le J_n} (D_K^*(\langle v_j, U_j\rangle))^{1/2}>d_n\} = O(\sum_n J_n n^{-2+\varepsilon})<\infty.$$

Therefore by the Borel-Cantelli lemma,

$$(4.19) \qquad \max_{1\le j\le J} (D_K^*(\langle v_j, U_j\rangle))^{1/2} = O(d_n), \qquad n\to\infty,$$

almost surely. The proof now follows by combining (4.2) and (4.19), with the stated choice of J_n. //

For $p = 1$ the rate of convergence in (4.15) is $(\log n)^{1/4} n^{-1/6}$, which, while slower than that of (3.11) for $f \in$ Lip 1, exceeds the latter rate for less smooth f and even applies to discontinuous f of bounded variation.

5. Alternative Methods of Estimation in L^∞. For L^∞-estimation of continuous f, there are other ways to use the data $\{(t_{jk}, s_{jk})\}$ to estimate f. Two such methods are developed in this section. The first involves an estimator of rather different form than the $\hat{f}$ given by (2.5), and for it we obtain an error bound less than that in (3.1) but of the same order of magnitude. The second method uses an adaptive choice of the u_{jk} for each j and achieves a dramatically improved rate of convergence: n^{-1} for all f satisfying any Lipschitz condition.

The methods of this section are based on the observation that the value of s_{jk} determines whether the point $(t_{jk}, -2Bu_{jk}+B)$ is below, on or above the graph of f, and that given two such points on opposite sides of the graph, the line segment joining them intersects the graph of f (by continuity of f).

Since discrepancy does not enter the analysis, it is convenient to replace the u_{jk} by points $x_{jk} \in [-B, B]$, where B is the bound on $|f|$. The data, therefore, is represented as

(5.1a) $$\{(t_{jk}, x_{jk}, b_{jk}): 1 \le j \le J,\ 1 \le k \le K\},$$

where the t_{jk} satisfy (2.3b), the x_{jk} are in $[-B,B]$, and

(5.1b) $$b_{jk} = \mathrm{sgn}\,(f(t_{jk})-x_{jk}).$$

Note that whereas in Section 2 the u_{jk} were not treated as data

(since their values are not necessary to compute the estimator $\hat{f}$, although they do appear in the bound (3.1)) the x_{jk} now must be part of the data since their values are needed to calculate the estimator $\hat{f}$ defined in (5.2) below. We assume, still, that there are intervals $I_j = [t_{j-1}, t_j)$ satisfying $t_{jk} \in I_j$ for each j and k and that the estimator $\hat{f}$ is to be a step function satisfying (2.4) whose value on I_j depends only on $\{(t_{jk}, x_{jk}, b_{jk}): 1 \le k \le K\}$.

For the first estimator the x_{jk} are fixed in advance, i.e., the estimator is nonadaptive. Its value on I_j is determined as follows. Let $-B = x_{j,(0)} \le x_{j,(1)} \le \cdots \le x_{j,(K)} \le x_{j,(K+1)} = B$ be the order statistics of the points $-B, x_{j1}, \ldots, x_{jK}, B$, let σ be that permutation of $\{1,\ldots, K\}$ for which $x_{j,\sigma(i)} = x_{j,(i)}$ for each i, and let Γ denote the polygonal path with vertices $(t_{j-1}, x_{j,(0)}), (t_{j,\sigma(1)}, x_{j,(1)}), \ldots, (t_{j,\sigma(K)}, x_{j,(K)})$ and $(t_j, x_{j,(K+1)})$ in that order (these are merely the points $(t_{j-1}, -B), (t_{j1}, x_{j1}), \ldots, (t_{jK}, x_{jK}), (t_j, B)$ with the second coordinates in increasing order). Since f is continuous at least one segment of Γ intersects the graph of f and which segments do so can be determined from the data (5.1). Choose any such segment, say that with endpoints $(t_{j,\sigma(i)}, x_{j,(i)})$ and $(t_{j,\sigma(i+1)}, x_{j,(i+1)})$, and define

$$\hat{f}(t) = \tfrac{1}{2}(x_{j,(i+1)} + x_{j,(i)}), \qquad t \in I_j. \tag{5.2}$$

The following result provides an error bound for *this* estimator.

THEOREM 5.1. Assume that f is continuous and that $|f| \le B$ on $[0,1]$ and let $\hat{f}$ be given by (5.2). Then

$$\|\hat{f} - f\|_\infty \le \tfrac{1}{2} \max_{1\le j\le J} \max_{1\le k\le K+1} (x_{j,(k)} - x_{j,(k-1)}) + \max_{1\le j\le J} \omega(f;\, t_j - t_{j-1}). \tag{5.3}$$

(Recall that $\omega(f;.)$ is the modulus of continuity of f.)

PROOF. Fix j. If $\hat{f}(t), t \in I_j$, is given by (5.2), then in the interval with endpoints $t_{j,\sigma(i)}$ and $t_{j,\sigma(i+1)}$ there is $\tilde{t}$ such that $f(\tilde{t})$ lies on Γ. Hence for $t \in I_j$,

$$|\hat{f}(t) - f(t)| \le |\hat{f}(t) - f(\tilde{t})| + |f(\tilde{t}) - f(t)|$$
$$\le \tfrac{1}{2}(x_{j,(i+1)} - x_{j,(i)}) + \omega(f;\, t_j - t_{j-1}),$$

and (5.3) follows immediately. //

With J and K fixed the right-hand side of (5.3) is minimized by choosing $t_j = j/J$ and the x_{jk} such that $x_{j,(i)} - x_{j,(i-1)} = 2B(K+1)^{-1}$. This transforms (5.3) to

$$\|\hat{f} - f\|_\infty \le \frac{B}{K+1} + \omega(f;\, \tfrac{1}{J}), \tag{5.4}$$

which slightly improves (3.6) for finite K and is equivalent in order as $K\to\infty$. *If the* x_{jk} *are increasing in* k *for each* j, *then* the computational and storage requirements for (2.5) and (5.2) are comparable. However, whereas (2.5) can be updated recursively as K increases, (5.2) cannot.

COROLLARY 5.1. If $f \in \mathrm{Lip}\ \gamma$ satisfies the assumptions of Theorem 5.1 and $\hat{f}$ is given by (5.2) with $J = J_n = n^{1/(\gamma+1)}$ and $t_j = j/J$, and with $K=K_n=n^{\gamma/(\gamma+1)}$ and $x_{jk} = -B + kB/(K+1)$ for each j and k, then

$$\|\hat{f} - f\|_\infty = O(n^{-\gamma/(\gamma+1)}), \qquad n\to\infty. \tag{5.5}$$

Using the easily established inequality

$$\max_{1\le k\le K} (x_{j,(k)} - x_{j,(k+1)}) \le 4B\, D_K^*(\langle u_j \rangle),$$

where $x_{jk} = 2Bu_{jk} - B$, one can obtain an *almost sure convergence rate* for the case where the x_{jk} are replaced by independent, *uniform* $[-B,B]$ random variables X_{jk}; the rate is that of (3.11). We omit further details.

Although seemingly unrelated, the methods yielding the estimators (2.5) and (5.2) can be viewed as two approaches to the same numerical integration problem. For $y \in [-B,B)$, recall that $y = \int_0^1 q(y,u)du$. The method of Sections 2 and 3 approximates the integral - within the restrictions we have imposed - as an average of q-

values. However, for fixed y, the function $u \to q(y,u)$ assumes only the known values $-B$, 0 and B so that its integral can be estimated from an estimate of its single point of discontinuity, namely $u^*(y) = (B-y)/2B$. The estimator $\hat{f}$ in (5.2) estimates this point of discontinuity from the data (5.1).

Using the <u>nonadaptive</u> estimator $\hat{f}$ of (5.2) one can "pin down" the value of f somewhere in the interval I_j only to within

$\max_k (x_{j,(k)} - x_{j,(k-1)}) \cong 2BK^{-1}$. By choosing the x_{jk} <u>adaptively</u>, one can do much better and can obtain rates of convergence $\|\hat{f} - f\|_\infty = O((\log n)n^{-1})$ for all f satisfying any Lipschitz condition. We now describe the algorithm for constructing $\hat{f}$, which is based on the data (5.1), except that now the x_{jk} will be determined adaptively – but recursively – for each j. Assume that $|f| \leq B$ on $[0,1]$. The value of $\hat{f}$ on I_j is constructed as follows.

<u>Step 1 (Initialization)</u>. Note that the points (t_{j1},B) and $(t_{j1},-B)$ are above and below the graph of f, respectively. Set $(t_0^+,x_0^+) = (t_{j1},B)$, $(t_0^-,x_0^-) = (t_{j1},-B)$.

<u>Step 2 (Iteration)</u>. The k^{th} step of the iteration is entered with two points (t_{k-1}^+,x_{k-1}^+) and (t_{k-1}^-,x_{k-1}^-), from among the points $(t_{j1},x_{j1}),\ldots,(t_{j,k-1}, x_{j,k-1})$, (t_{j1},B) and $(t_{j1},-B)$, such that (t_{k-1}^+, x_{k-1}^+) is above the graph of f, (t_{k-1}^-,x_{k-1}^-) is below the graph of f, and $|x_{k-1}^+ - x_{k-1}^-| = B2^{-k+2}$. Then set $x_{jk} = \frac{1}{2}(x_{k-1}^+ + x_{k-1}^-)$ and calculate $b_{jk} = \mathrm{sgn}\,(f(t_{jk}) - x_{jk})$.

a) If $b_{jk} = 0$, then $f(t_{jk}) = x_{jk}$; proceed to the termination step.

b) If $b_{jk} = 1$, then (t_{jk},x_{jk}) is below the graph of f. Let $(t_k^+,x_k^+) = (t_{k-1}^+,x_{k-1}^+)$ and $(t_k^-,x_k^-) = (t_{jk},x_{jk})$, and proceed to the next iteration.

c) If $b_{jk} = -1$, then (t_{jk},x_{jk}) is above the graph of f. Let $(t_k^+,x_k^+) = (t_{jk},x_{jk})$ and $(t_k^-,x_k^-) = (t_{k-1}^-,x_{k-1}^-)$, and proceed to the next iteration.

Note that for either b) or c) we have $|x_k^+ - x_k^-| = \frac{1}{2}|x_{k-1}^+ - x_{k-1}^-|$.

<u>Step 3 (Termination)</u>. There are two possibilities.

a) If there is k such that $f(t_{jk}) = x_{jk}$, set

$$\hat{f}(t) = x_{jk}, \qquad t \in I_j. \tag{5.6a}$$

b) Otherwise, the procedure terminates with evaluation of b_{jK} and yields points (t_K^+,x_K^+) above the graph of f and (t_K^-,x_K^-) below the graph of f such that $|x_K^+ - x_K^-| = B2^{-K+1}$. In this case set

$$\hat{f}(t) = \tfrac{1}{2}(x_K^+ + x_K^-), \qquad t \in I_j. \tag{5.6b}$$

Note that the t_{jk} are not chosen adaptively, but can be specified in advance. Furthermore, the algorithm is recursive in that only the current values of (t_k^+,x_k^+) and (t_k^-,x_k^-) need be stored in order to determine either $x_{j,k+1}$ or the value of the estimator. The resulting error bound dramatically improves those for the nonadaptive estimators (2.5) and (5.2).

THEOREM 5.2. <u>Assume that</u> f <u>is continuous and that</u> $|f| \leq B$ <u>on</u> $[0,1]$ <u>and let</u> $\hat{f}$ <u>be given by</u> (5.6). <u>Then</u>

$$\|\hat{f} - f\|_\infty \leq B2^{-K} + \max_{1 \leq j \leq J} \omega(f; t_j - t_{j-1}). \tag{5.7}$$

PROOF. Let j be fixed and consider the two possible forms of termination separately.

a) If $f(t_{jk}) = x_{jk}$ and (5.6a) holds, then for $t \in I_j$,

$$|\hat{f}(t) - f(t)| = |f(t_{jk}) - f(t)| \le \omega(f; t_j - t_{j-1}).$$

b) If $\hat{f}$ is given by (5.6b) then by continuity of f there is $\bar{t}$ within the interval with endpoints t_K^+ and t_K^- such that $f(\bar{t})$ is on the line segment joining (t_K^+, x_K^+) and (t_K^-, x_K^-). *Hence, for* $t \in I_j$,

$$\begin{aligned} |\hat{f}(t) - f(t)| &\le |\tfrac{1}{2}(x_K^+ + x_K^-) - f(\bar{t})| + |f(\bar{t}) - f(t)| \\ &\le \tfrac{1}{2}|x_K^+ - x_K^-| + \omega(f; t_j - t_{j-1}) \\ &= B2^{-K} + \omega(f; t_j - t_{j-1}). \end{aligned}$$

Consequently, (5.7) holds. //

For fixed J, the right-hand side of (5.7) is minimized for $t_j = j/J$, yielding the bound

(5.8) $$\|\hat{f} - f\|_\infty \le B2^{-K} + \omega(f; \tfrac{1}{J}).$$

COROLLARY 5.2. Assume that $f \in \mathrm{Lip}\,\gamma$ and that $|f| \le B$ on $[0,1]$. Let $\hat{f}$ be given by (5.6) with $K = K_n$ satisfying $K(2^{1/\gamma})^K = n$, $J = J_n = n/K$ and $t_j = j/J$ for each j. Then

(5.9) $$\|\hat{f} - f\|_\infty = O((\log n)n^{-1}), \qquad n \to \infty.$$

PROOF. From (5.8), there is a constant C such that

$$\|\hat{f} - f\|_\infty \le C\left(\frac{1}{2^{1/\gamma}}\right)^K = C(\frac{K}{n}) \le C\frac{\log_\alpha n}{n}$$

$$= O(\frac{\log n}{n}),$$

where $\alpha = 2^{1/\gamma}$. //

The algorithm leading to (5.6) can be improved in practice, although not in the worst case (i.e., the bound (5.7) is not improved), by the following device. If at any iteration the points (t_k^+, x_k^+) above the graph of f and (t_k^-, x_k^-) below satisfy $x_k^+ \le x_k^-$, which is easily checked, then somewhere in the interval with endpoints t_k^+, t_k^-, f must assume the value $\frac{1}{2}(x_k^+ + x_k^-)$. If one takes $\hat{f}(t)$ to be this value, then the first term on the right-hand side of (5.7) is unnecessary and one has $|\hat{f} - f| \le \omega(f; t_j - t_{j-1})$ on I_j.

6. Complements. In this section we sketch an extension of Theorem 3.1 to the case of functions f defined on $[0,1]^d$ for some $d \ge 2$. In addition, we include a few comments concerning our method of reconstruction.

We first consider reconstruction of functions on $[0,1]^d$, $d \ge 2$. The estimator $\hat{f}$ is constructed in the following manner. Partition $[0,1]$ into intervals I_j as in Section 2 and let the t_{jk} satisfy (2.3b). Suppose that for each choice of $\underset{\sim}{j} = (j_1, \ldots, j_d)$, where $0 \le j_p \le J$ for each j, there are K^d numbers $\{(u_{\underset{\sim}{jk}}): 1 \le k_r \le K\}$ in $[0,1]$. By analogy with (2.5) we introduce the estimator

(6.1) $$\hat{f}(\underset{\sim}{t}) = \frac{1}{K^d}\sum_{\underset{\sim}{k}} q(f(t_{j_1k_1}, \ldots, t_{j_dk_d}), u_{\underset{\sim}{jk}})), \quad \underset{\sim}{t} \in I_{j_1} \times \cdots \times I_{j_d}.$$

The work (i.e., number of function evaluations) required to calculate $\hat{f}$ is $n = (JK)^d$.

An essentially verbatim repetition of the proof of Theorem 3.1 yields

THEOREM 6.1. Assume that $|f| \le B$ on $[0,1]^d$ and that $\hat{f}$ is defined by (6.1). Then

$$\|\hat{f} - f\|_\infty \le 2B \max_j D^*_{K^d}(\langle u_j \rangle) + \max_j \omega(f; I_{j_1} \times \dots \times I_{j_d}), \tag{6.2}$$

where $D^*_{K^d}(\langle u_j \rangle)$ is the discrepancy (in $[0,1]^d$) of the numbers $\{(u_{jk}): 1 \le k_r \le K\}$ and where $\omega(f;A)$ is the oscillation function of f, given by

$$\omega(f; A) = \sup\{|f(t) - f(s)|: t, s \in A\}.$$

By suitable choice of the u_{jk} - see the comments following *Theorem* 3.1 - and for equally spaced t_j, it is easily verified that if f satisfies

$$|f(t) - f(s)| \le \|t - s\|^\gamma$$

for some $\gamma \in (0,1]$, where $\|t - s\|$ is the Euclidean norm of $t - s$, then with an optimal *division* of labor between integration and interpolation,

$$\|\hat{f} - f\|_\infty = O(n^{-\gamma/(d+\gamma)}), \qquad n \to \infty, \tag{6.3}$$

where $n = (JK)^d$ is the work required to calculate $\hat{f}$ using (6.1).

The results in Section 4 on L^p convergence extend analogously.

We conclude with some comments concerning our method for reconstructing a function. First, our bounds for errors of the *form* $\|\hat{f} - f\|_\infty$ are insensitive to the distribution of the points t_{jk} in each interval I_j. It is possible that in some cases *one could have*, if desired, $t_{j1} = \dots = t_{jK}$. (However, in many applications this will be impossible because of the sequential nature of the data collection procedure.) Being able to evaluate $q(f(t_j^*), u_{jk})$ for K values of u_{jk} in effect permits discretization of the function f at t_j^*. If the u_{jk} are increasing in k, this is precisely what the estimator (5.2) accomplishes. One cannot improve the bound (5.3) by taking the t_{jk} to be equal for fixed j, nor can any other of our L^∞ bounds be similarly improved. Regarded in this context, our estimator (2.5) is as effective as the estimators obtained by discretizing f at one point in each interval I_j.

Here are two final points. In practice the bound B on f may not be known; our estimators then serve simply to estimate the truncated function f_B defined by

$$f_B(t) = \begin{cases} B & f(t) > B \\ f(t) & -B \le f(t) \le B \\ -B & f(t) < -B. \end{cases}$$

Also, the restriction to $[0,1]$ as the domain of f is inessential; one can replace $[0,1]$ by *any finite interval* $[a,b]$, although the constants in some of the bounds will be multiplied by b-a. Since our estimators are local on the intervals I_j, estimation over, for example, $[0,\infty)$ is even possible.

In effect, our procedure estimates f based on observations through a window of the form $[0,1] \times [-B,B]$, which can be replaced by any other rectangular window.

__Acknowledgement__. We are indebted to our colleague Richard H. Byrd for some discussions stimulating the development of Section 5.

References

[1] H. Berens and R. A. DeVore, *Quantitative Korovkin theorems for L_p - spaces*, Approximation Theory II, G. G. Lorentz, C. K. Chin, L. L. Schumaker, eds., Academic Press, New York, 1976, pp. 289-298.

[2] K. L. Chung, *An estimate concerning the Kolmogorov limit distribution*, Trans. Amer. Math Soc. 67 (1949) 36-50.

[3] A. Dvoretsky, J. Kiefer and J. Wolfowitz, *Asymptotic minimax character of the sample distribution function and of the classical multinomial estimator*, Ann. Math. Statist. 27 (1956) 642-669.

[4] J. Kiefer, *On large deviations of the empiric D.F. of vector chance variables and a law of iterated logarithm*, Pacific J. Math. 11 (1961) 649-660.

[5] J. F. Koksma, *Een algemeene stelling uit de theorie der gelijkmatige verdeeling modulo 1*, Mathematica B (Zutphen) 11 (1942/43) 7-11.

[6] L. Kuipers and H. Niederreiter, *Uniform Distribution of Sequences*, Wiley-Interscience, New York, 1974.

[7] E. Masry, *The reconstruction of analog signals from the sign of their noisy samples*. IEEE Trans. Inf. Theory (to appear).

[8] E. Masry and S. Cambanis, *Signal identification after noisy nonlinear transformations*. IEEE Trans. Inf. Theory IT-26 (1980) 50-58.

[9] E. Masry and S. Cambanis, *Consistent estimation of continuous-time signals from nonlinear transformations of noisy samples*. IEEE Trans. Inf. Theory IT-27 (1981) 84-96.

[10] H. Niederreiter, *Quasi-Monte Carlo methods and pseudo-random numbers*, Bull. Amer. Math. Soc. 84 (1978) 957-1041.

[11] A. F. Timan, *Theory of Approximation of Functions of a Real Variable*, Macmillan, New York, 1963.

Unclassified
SECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE

1. REPORT NUMBER ONR No. 81-4	2. GOVT ACCESSION NO. AD-A104	3. RECIPIENT CATALOG NUMBER 477
4. TITLE ERROR BOUNDS FOR RECONSTRUCTION OF A FUNCTION f FROM A FINITE SEQUENCE $\{SGN(f(t_i) + x_i)\}$		5. TYPE OF REPORT & PERIOD COVERED Technical Report 6. PERFORMING ORGANIZATION REPORT NO. Technical Report No. 341
7. AUTHOR(s) Alan F. Karr and Robert J. Serfling		8. CONTRACT OR GRANT NUMBER(s) ONR No. N00014-79-C-0801
9. PERFORMING ORGANIZATION NAME AND ADDRESS Department of Mathematical Sciences The Johns Hopkins University Baltimore, Maryland 21218		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME & ADDRESS Office of Naval Research Statistics and Probability Program Arlington, Virginia 22217		12. REPORT DATE May, 1981 13. NUMBER OF PAGES 36
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)		15. SECURITY CLASS (of this report) *Unclassified* 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS

nonparametric signal identification; approximation; convergence rates.

20. ABSTRACT Consider reconstructing a function $f(t)$, $0 \le t \le 1$, from knowledge only of $\{(t_i,s_i), 1 \le i \le n\}$, where $s_i = \mathrm{sgn}(f(t_i) + x_i)$, $1 \le i \le n$, and the x_i are additive "corruptions." Without the x_i, f could not be reconstructed. However, for f continuous and for random uniform $\bar{x}$, Masry and Cambanis (1980,1981) show that f can be consistently estimated almost surely and in mean square as $n \to \infty$. In the present treatment the approximation of $f(t_i)$ is identified as a numerical integration problem rather than a statistical problem. We obtain simple bounds on the error of estimation, allow arbitrary (random or deterministic) noise x, and deal with the case of discontinuous f. The bounds yield substantially improved convergence rates for x_i a quasi-random rather than random sequence.